Stochastic and doubly stochastic spatio-temporal field modeling for data assimilation

Michael Tsyrulnikov, Dmitry Gayfulin and Alexander Rakitko

HydroMetCentre of Russia

Melbourne, 5 December 2016
Outline

1. **A limited area Stochastic Pattern Generator (SPG) for model error simulation**
   - Model errors: a brief introduction
   - The SPG

2. **Doubly stochastic models** for simulation of “truth” in testing DA methodologies
   - A one-variable model
   - A stochastic advection-diffusion-decay model on the circle
Model-error modeling
Model (tendency) errors

Consider a forecast equation:

\[ \frac{dx}{dt} = F(x) \] (1)

Due to imperfections in \( F \), the truth satisfies

\[ \frac{dx}{dt} = F(x) - \xi \] (2)

where \( \xi \) is the model error.
Model error models

1. Non-stochastic
   - Multi-model
   - Multi-physics
   - Multi-parameter

2. Stochastic
   - Additive perturbations, like Stochastic Kinetic Energy Backscatter (SKEB).
   - Multiplicative perturbations, like Stochastic Perturbations of Physics Tendencies (SPPT)
   - Stochastically Perturbed Parameterizations (SPP)

All the existing stochastic model-error models require a stochastic spatio-temporal pattern generator.
The SPG: motivation
Space-time interactions and proportionality of scales

The existing **pattern generators** for ensemble applications produce *separable* space-time correlations:

\[ C(t, s) = C_t(t) \cdot C_s(s) \]

But: no space-time interactions.

In reality, larger spatial scales ‘live longer’ than smaller spatial scales, which ‘die out’ quicker. For large \( k \),

\[ \tau_k \sim \frac{1}{k} \]

This ‘proportionality of scales’ is widespread in geophysical fields (Tsyroulnikov QJRMS 2001) and other media.
Illustration:
Separable vs. Proportional-scales model error models

**Experiment**

- Simulate model error fields on a 1D spatial domain with:
  - Separable correlations: \( \exp(-|\Delta x|/L) \cdot \exp(-U|\Delta t|/L) \)
  - Proportional-scales correlations: \( \exp(-\sqrt{(|\Delta x|^2 + (U\Delta t)^2)/L}) \)

Note that both fields have exactly the same spatial length scales and exactly the same temporal length scales.

**Setup:**

The domain: 100 km \( \times \) 3 h.
The grid: 100 \( \times \) 100 points.
\( L_{05} = 50 \text{ km} \) and \( U = 20 \text{ m/s} \).
Separable vs. Proportional-scales fields: model errors

Model error: Separable

Model error: Proportional Scales
Illustration: impact on forecast errors

For small model errors and small time periods, the forecast model operator $F$ can be linearized, so that the forecast error $\delta x$ due to the model error $\xi$ satisfies

$$\frac{d\delta x}{dt} = A\delta x - \xi$$

so that

$$\delta x(t) = e^{At} \int_0^t e^{-A\tau} \xi(\tau) \, d\tau = \int_0^t \xi(\tau) \, d\tau + o(t^2) \approx \int_0^t \xi(\tau) \, d\tau$$

- Integrate both fields in time, getting, approximately, forecast errors.
- Examine the spatial length scales of the forecast errors as functions of time.
Separable vs. Proportional-scales fields: forecast errors

Model error field

Time integrated model error

Michael Tsyrulnikov, Dmitry Gayfulin and Alexander Rakitko (HMC)
Dynamical relevance of space-time interactions:

- **Separability** $\Rightarrow$ the **spatial length scale** $L_x$ of the **forecast error** field is **constant**.

- **Proportionality of scales** $\Rightarrow$ $L_x$ **grows** with $t$. $$L_x = \frac{\text{s.d. } \delta}{\text{s.d. } \partial \delta / \partial x}$$

**NB:** The forecast errors length scale is a very important attribute of the background-error covariances.
The SPG: design
Formulation of the SPG

**Requirement:** The generated fields should have “proportional scales”.

**Approach:** Linear evolutionary stochastic partial differential equations.

**Starting point:**

\[
\frac{\partial \xi(t, s)}{\partial t} + A \xi(t, s) = \alpha(t, s)
\]

= \( t \) is time, \( s = (x, y, z) \) is the spatial vector
= \( \alpha \) is the white driving noise
= \( A \) is the spatial operator.
= \( \xi \) is the output random field

- To get the spatial isotropy, we specify \( A = P(-\Delta) = \mu(1 - \lambda^2 \Delta)^q \), where \( \Delta \) is the spatial Laplacian and \( P \) is the polynomial or order \( q \).
- We impose the “proportionality of scales’ (\( \tau_k \sim 1/k \) as \( k \to \infty \)) \Rightarrow

\[
q = \frac{1}{2}
\]
Formulation of the SPG

Model:

\[
\left( \frac{\partial}{\partial t} + \mu \sqrt{1 - \lambda^2 \Delta} \right)^3 \xi(t, s) = \sigma \alpha(t, s)
\]

\(\sigma\) controls the variance
\(\lambda\) controls the spatial scale
\(\mu\) controls the temporal scale

Computational domain: the cube with cyclic boundary conditions.
Numerical scheme: spectral in space and finite-difference in time.
Spatio-temporal covariances

Ranges: $t=0...12$ h, $r=0...750$ km
Spatial field (xy)

Horizontal cross-section of the random field

---

Michael Tsyrlulnikov, Dmitry Gayfulin and Alexander Rakitko (HMC)
Stochastic and doubly stochastic spatio-temporal field modeling for data assimilation
Melbourne, 5 December 2016
Space-time plot (xt)

Spatio-temporal cross-section of the random field
Application with the COSMO model

3h V-wind forecast perturbation. Model level 22
Conclusions on the SPG

The SPG produces Gaussian pseudo-random fields on a 2D or 3D limited area domain with non-separable spatio-temporal correlations.

= The Fortran code of the SPG is freely available from github.com/gayfulin/SPG.

= Tsyrulnikov M. and Gayfulin D. A limited-area spatio-temporal stochastic pattern generator for ensemble prediction and ensemble data assimilation. – Meteorol. Zeitschrift, 2016 (accepted). (A copy can be downloaded from ResearchGate.)
Doubly stochastic models of “truth”
Simple models of “truth”

- Goal: in-depth investigation of DA methodologies in a simplified setting with synthetic and thus known “truth”.
- Examples: nonlinear one-variable models (the logistic map, the Henon map) and multi-variable models (the Lorenz models, the Burgers’ equation on the circle).
- A step forward: we wish to know not only the true state but also the true statistics.
- Bishop and Satterfield (MWR 2013, Part 1) computed the true EnKF error statistics with a deterministic model of truth and a stochastic forecast model.
- In our doubly stochastic models, the truth is stochastic whereas the forecast model is deterministic.
Building a 1D doubly stochastic model of “truth”

We start with the simplest time discrete stochastic equation

\[ x_k = F_k x_{k-1} - \sigma_k \varepsilon_k \]

If \( F_k = \text{const} \) and \( \sigma_k = \text{const} \), then the solution \( x_k \) is a stationary random process.

In reality, the dynamics and statistics of the truth vary in time (and space).
To mimic this variability, we specify $F_k$ and $\sigma_k$ to be random processes by themselves:

$$F_k = \bar{F} + \mu(F_{k-1} - \bar{F}) + \sigma_F \varepsilon_k$$

$$\sigma_k = \exp(\Sigma_k)$$

$$\Sigma_k = \kappa \Sigma_{k-1} + \sigma \Sigma \varepsilon_k$$

$$x_k = F_k x_{k-1} + \sigma_k \varepsilon_k$$
The doubly stochastic advection-diffusion-decay model

**Non-stochastic:**
\[
\frac{\partial x}{\partial t} + U \frac{\partial x}{\partial s} + \rho x - \nu \frac{\partial^2 x}{\partial s^2} = 0
\]

(*t* is time, *s* is the spatial coordinate, *U* is the advection velocity, *\rho* is the decay coefficient, and *\nu* is the diffusion coefficient).

**Singly stochastic:**
\[
\frac{\partial x}{\partial t} + U \frac{\partial x}{\partial s} + \rho x - \nu \frac{\partial^2 x}{\partial s^2} = \sigma \cdot \alpha(t, s)
\]  

(*\alpha* is the driving noise and *\sigma* its intensity).

**Doubly stochastic:**
\[
\frac{\partial x}{\partial t} + U(t, s) \frac{\partial x}{\partial s} + \rho(t, s) x - \nu(t, s) \frac{\partial^2 x}{\partial s^2} = e^{\Sigma(t,s)} \alpha(t, s)
\]

where the coefficients,
*U(t, s)*, *\rho(t, s)*, *\nu(t, s)*, and *\Sigma(t, s)* (or some of them),
are **spatio-temporal random fields** by themselves postulated to satisfy the singly stochastic advection-diffusion-decay model (*\alpha*).
A space-time $x(t, s)$ plot for our model of truth

sd$_U$ = 30
A space-time $x(t, s)$ plot for the Lorenz-96 model

$$\text{dim} = 32, \ F = 8$$
Hierarchical simulation

\[ x_k = F_k x_{k-1} + \sigma_k \varepsilon_k \]

1. Specify parameters of the processes \( F_k \) and \( \sigma_k \). This defines a “universe”. All worlds in a “universe” share the same “climate”.

2. Generate \( \{F_k\} \) and \( \{\sigma_k\} \). This pair of sequences defines a “galaxy”. All worlds in a “galaxy” share the same “statistics of the day”, e.g. \( \text{Var} \ x_k \) for each \( k \).

3. Given the sequences \( \{F_k\} \) and \( \{\sigma_k\} \), generate \( \{\varepsilon_k\} \) and thus \( \{x_k\} \) getting “worlds” in the “galaxy”. \( \text{Var} \ x_k \) can be estimated by averaging over worlds within the galaxy. In the same way, true error covariances, say \( B_k \), can be estimated for any filter in question.
Space-time plots for the variance of the “truth” $\text{Var } x(t, s)$

The 4 plots correspond to the 4 different variances of the fields $\rho(t, s)$ and $\Sigma(t, s)$.
Examples of applications of the doubly stochastic models of truth
Stochastic 1D EnKF with the optimally tuned inflation 1.01
Stochastic 64-D EnKF:

Errors in (localized) sample correlations, $S_{i,i+2} - B_{i,i+2}$, vs. true $B_{i,i+2}$
Conclusions on the doubly stochastic models of “truth”

- Are capable of generating complicated spatio-temporal variability.
- Provide the “true” time and space specific spatio-temporal error statistics (instantaneous variances, length scales, etc.)
- Allow the use of the exact KF (as a benchmark).

Thank you!